UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

CRT-601

A SELF ADJUSTING INDEX FOR CATEGORICALLY ORDERED DATA*

by

Marc Gaudry[1]
and
Richard Laferrière[2]

[1] Département de Sciences économiques and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, Succursale A, Montréal, H3C 3J7.

[2] Economics Department, Concordia University, 1455 De Maisonneuve Blvd. West, H3G 1M8, and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, Succursale A, Montréal, H3C 3J7.

* The research on which this paper is based was supported at various stages by Transport Canada, the F.C.A.R. program of Québec, the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank Denyse L. Dagenais and Michael Strobel for useful discussions.

Université de Montréal
Centre de recherche sur les transports - Publication #601
Département de Sciences économiques - Cahier #8835

Concordia University
Economics Department - Publication #1988-11

October 1988

Dépôt légal - 4e trimestre 1988

ABSTRACT

We propose a new index to treat information obtained from questionnaires with categorically ordered answers. We also show how the index is useful if ordinal answers are cardinalized by the researcher, and provide an example of such a case.

Key words: QUESTIONNAIRE, CATEGORICALLY ORDERED DATA, BOX-COX TRANSFORMATION, LATENT VARIABLES

RESUME

Nous proposons un nouvel indice pour traiter les réponses ordonnées par catégories. Nous montrons la pertinence de l'indice dans le cas de réponses ordinales cardinalisées par l'analyste, et présentons un tel exemple.

Mots-clés: TRANSFORMATION BOX-COX, QUESTIONNAIRE, REPONSES ORDONNEES PAR CATEGORIES, VARIABLES LATENTES.

CONTENTS

1. Introduction....3
2. The SACOD index and its properties....3
   2.1 Data scale compensation....4
   2.2 Indifference to shift of analyst's values....5
   2.3 Indifference to location of the origin....6
   2.4 Endogenous curvature and relative importance of observations....6
3. Possible uses....8
   3.1 Extension of Van Praag's formulation....8
   3.2 Usefulness with uncardinalized values....10
      3.2.1 Length of the domain....10
      3.2.2 Inaccuracy of the relative position of the answers....11
      3.2.3 Example: a grieving reaction index....12
         A. Illustrative model....12
         B. Self adjustment and sensitivity to cardinalization....17
4. Conclusion....18
References....18

Figure 1 Curvature of the SACOD index....7
Table 1 Determinants of grieving reaction: self adjustment and error sensitivity of the index....14

A SELF ADJUSTING INDEX FOR CATEGORICALLY ORDERED DATA*

1. INTRODUCTION

This article introduces a new index to treat information obtained from questionnaires with categorically ordered answers. After a presentation of the new index and its properties under the assumption that questionnaire answers are cardinal, we show how our index could extend the individual welfare function measurement approach originated by Van Praag (1968). We also show how the index is useful if ordinal answers are cardinalized by the researcher, and provide an example of such a case.

For situations in which one is trying to circumscribe a phenomenon from many vantage points, the cardinalization of ordinal answers can be seen as an alternative to the use of the ordered LOGIT and LISREL techniques. The ordered logit, a privileged technique for the treatment of successive categories rating data, requires as many models as there are questions in the questionnaire. LISREL rapidly leads to overidentification problems when many observed measures of the latent phenomenon are supplied.

2. THE SACOD INDEX AND ITS PROPERTIES

Often, individuals are asked their opinion on the performance of a government, or on their well-being, and producers are questioned about their incentive to invest, etc. For instance, a typical question of such questionnaires might be: *Express the strength of your willingness to give blood on a scale from 1 to 8.* The data obtained from such questionnaires are completely characterized by a domain, i.e. a numerical maximum and minimum, and the relative positions implied by the actual values of the observations.

The questionnaire might contain many similar questions aimed at obtaining the same information from many facets and isolating a latent phenomenon. To handle such cases, we define the following index:

$$S = \left[\left(\prod_{i=1}^{n} \exp(R_i)/\exp(R_{max}) \right)^{1/n} \right]^{(\lambda)}, \qquad (1\text{-A})$$

or

$$S = \left[\exp\left(\sum_{i=1}^{n} (R_i - R_{max})/n \right) \right]^{(\lambda)}, \qquad (1\text{-B})$$

---

* The research on which this paper is based was supported at various stages by Transport Canada, the F.C.A.R. program of Québec, the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank Denyse L. Dagenais and Michael Strobel for useful discussions.

where R_i is the numerical value of the answer given to the i^{th} of n questions and R_{max} the maximum possible value, 8 in the example below.[1] Expression (1-A) shows that these values are used to define a proportion, given by a ratio of exponentials, for each question, and that the geometric mean of these proportions is subjected to a Box-Cox transformation (BCT), i.e.

$$[\exp(\sum_{i=1}^{n}(R_i - R_{max})/n)]^{(\lambda)} = \{[\exp(\sum_{i=1}^{n}(R_i - R_{max})/n)]^{\lambda} - 1\}/\lambda, \text{ if } \lambda \neq 0,$$

$$= \sum_{i=1}^{n}(R_i - R_{max})/n, \qquad \text{if } \lambda = 0. \tag{2}$$

Note that, if $\lambda=0$, we obtain the arithmetic mean of deviations from the maximum score. In that special case, there is no interaction among the elements. In the general case, the Box-Cox transformation calibrates this interaction through a multidimensional adjustment of the data. For this reason, we call it the Self Adjusting Categorically Ordered Data (SACOD) index. We shall presently examine its properties and explicate the nature of the adjustments. All properties defined are valid when the number of questions is large, as well as in the single question case.

2.1 DATA SCALE COMPENSATION

The R_i's used in (1) come from the selection, by respondents, of values within a specified domain. This domain is suggested by the analyst, for instance 1 to 8 in our example above. We might then ask what would be the effect on the index of having used a different domain, say 10 to 80. Scaling the proposed domain by a constant b, we have

$$S = [\exp(\sum_{i=1}^{n}(b\,R_i - b\,R_{max})/n)]^{(\lambda)}$$

$$= b\,[\exp(\sum_{i=1}^{n}(R_i - R_{max})/n)^{b\lambda} - 1]/b\lambda$$

$$= b\,\{[\exp(\sum_{i=1}^{n}(R_i - R_{max})/n)]^{(\lambda^*)}\}, \tag{3}$$

where $\lambda^* = b\lambda$.

In a recent article, the present authors (Gaudry and Laferrière, 1988) have shown that the results obtained for the estimation of λ in (1) coincide with the results obtained for the estimation of λ^* in (3) because the λ^* parameter is unique. The unicity of λ^* implies that the calibrated index S adjusts the analyst's chosen domain by rescaling the data; conversely, had the analyst

---

1 For expository purposes and without loss of generality, we assume the same maximum and minimum values for every question.

changed the scale, which is tantamount to effecting a power transformation b, the resulting optimal λ^* would have been unchanged. Unicity also means that, when explanatory variables are used in a regression model with the index as dependent variable, their elasticities are unaffected by the analyst's chosen scale: only the regression coefficients are rescaled.

We note that, from a numerical point of view, the property of unicity also means that, were the scale of the $(R_i - R_{max})$ domain defined over an excessively large domain, an appropriate value of $\underline{b}$ would reequilibrate it (within the tolerance of the computer)[2].

2.2 INDIFFERENCE TO SHIFT OF ANALYST'S VALUES

Now consider another change in the analyst's presumed domain, namely shifting the domain by a specific constant $\underline{a}$; we then have

$$S = [\exp(\sum_{i=1}^{n} (a + R_i - R_{max} - a)/n)]^{(\lambda)}$$

$$= [\exp(\sum_{i=1}^{n} (R_i - R_{max})/n)]^{(\lambda)}. \qquad (4)$$

The fact that we can shift all observations by any constant without consequence is another flexible characteristic of the index.

The combination of this property with the data scale compensation property leads to the conclusion that one can approximate any desired domain D^* by a linear combination a + bD of the chosen domain D.

---

2 In Table 1, the reference model shown in Column 1 is specified with $\underline{b}$ arbitrarily set at .1 in order to be well within the precision of our computer, a CDC Cyber 835.

2.3 INDIFFERENCE TO LOCATION OF THE ORIGIN

The index has an upper bound B^u equal to 0 and an inferior bound B^i given by

$$B^i = [\ \exp(R_{min} - R_{max})\ ]^{(\lambda)}. \tag{5}$$

It can be transformed in such a way that $B^i = 0$ and B^u be determined by the λ parameter. This is the case since the origin of S, namely B^i, and therefore the benchmark R_{max} in (1), are completely arbitrary. To obtain $B^i = 0$, consider the scalar tranformation $s = \exp(R_{max} - R_{min})$; we then have

$$S^* = [s \exp(\sum_{i=1}^{n} (R_i - R_{max})/n\ )]^{(\lambda)},$$

$$= [\ \exp(\sum_{i=1}^{n} (R_i - R_{min})/n\ )]^{(\lambda)}. \tag{6}$$

Schlesselman (1971) has shown that the estimation of (λ) in (6) when S^* is the dependent variable of the model gives the same results as with S if, but only if, a constant is included among the regressors. However, if S^* is used instead of S, the *t*-statistics of the coefficients will not be invariant to the chosen origin since this choice necessitates a scaling of the untransformed dependent variable (Spitzer, 1984).

2.4 ENDOGENOUS CURVATURE AND RELATIVE IMPORTANCE OF OBSERVATIONS

Consider the average value of answers $\bar{R} \equiv \sum_{i=1}^{n} (R_i - R_{max})/n$. We clearly have

$$\frac{\partial^2 S}{\partial \bar{R}^2} = \lambda \exp(\bar{R})^{\lambda}. \tag{7}$$

In consequence, the index S is convex with respect to $\bar{R}$ if $\lambda > 0$, concave if $\lambda < 0$, and linear if $\lambda = 0$. The various cases are shown in Figure 1.

We note that, beyond a certain point, small values of $\bar{R}$ imply values of S that differ considerably from the values of S for $\lambda = 0$, where $\bar{R}$ just happens to be equal to S itself.

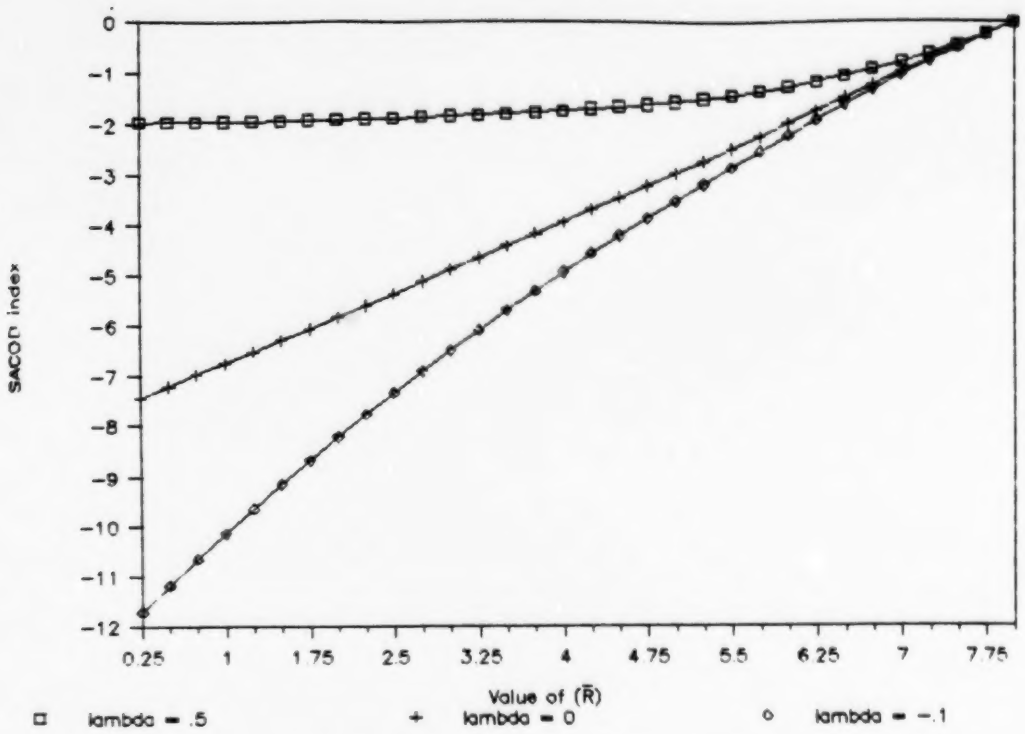

Figure 1 Curvature of the SACOD index

But curvature affects the relative position of observations. Suppose that n = 1 in (1) and let

$$S_k = \exp(R_k - R_{max})^{(\lambda)} \tag{8}$$

Then, consider the ratio of any two observations k and h:

$$\frac{S_k}{S_h} = \frac{\exp(R_k - R_{max})^{(\lambda)}}{\exp(R_h - R_{max})^{(\lambda)}} ; \tag{9}$$

it is clear that this ratio depends not only on the measured values for the k^{th} and h^{th} answers but also on the λ parameter. In general, the same holds for the relative values of any pair of answers considered in S.

In addition to the 4 characteristics just mentioned, one should state that our index easily accommodates missing data: they are automatically omitted in computing the geometric mean in (1-A). Note also that all available answers are used in the definition of the index: this obviates the need to select somehow only a subset of the answers through a priori or statistical devices.

3. POSSIBLE USES

3.1 EXTENSION OF VAN PRAAG'S FORMULATION

In a pioneering work, Van Praag (1968) introduced a new utility theory in order to measure an individual's welfare level with a function, called the Individual Welfare Function (IWF), from answers given to questionnaires. Our index allows an extension of that approach, which contains a number of steps.

Firstly, individuals are asked to associate a sum of money y_i to each of N ordered categories corresponding to qualitative rankings, such as excellent, good, sufficient, etc. These sums of money are then assumed to generate proportions of the satiety level assumed to be located at equal distances (i-1)/N on the individual's utility domain (that goes from 0 to 1). Van Praag (1968) justifies this equidistance rule by the assumption that individuals maximize information supplied.

Secondly, since utility is assumed to be lognormally distributed, it is possible to estimate each individual's σ and μ parameters of the function

$$U(y) = \frac{1}{\sigma^{1/2}(2\pi)^{n/2}} \int_{-\infty}^{\ln(y)} [\exp(-(t-\mu)'(t-\mu)/\sigma)]\, dt. \qquad (10)$$

where the values

$$U(y) = \{ 0, 1/N, 2/N, \ldots, (N-1)/N, 1 \} \qquad (11)$$

correspond to the sequence of declared amounts $y_1, y_2, \ldots, y_N$. The μ and σ parameters for each individual are estimated by doing the regression

$$\ln(y) = \mu + \sigma [Z^{-1}(U(y))] + \epsilon, \qquad (12)$$

where $Z^{-1}( )$ is the inverse normal distribution function. In the last step Van Praag explains the individuals' μ and σ estimates by socioeconomic factors.

We can extend this procedure in such a way that we obtain (11) as a special case. To do this, we propose to substitute

$$U^{*}(y) = \frac{\exp\left(\sum_{i=1}^{n} (y_i - y_{min})/n\right)^{(\lambda)}}{\exp\left(\sum_{i=1}^{n} (y_{max} - y_{min})/n\right)^{(\lambda)}}. \qquad (13)$$

for (11). This form is clearly a simple modification of the level format (1) effected in order to situate all utility points between 0 and 1, as is the case in (11) and the resulting share format naturally shares all the properties of the level format. Now it is easy to see that, for any value of λ in (13), there must be a set of declared y_i's that will yield the values assumed in (11): for instance, if $\lambda = 0$, the set $\{y_i = y_{i-1}+(y_{max}-y_{min})/N;$ $i=2,..,N\}$ will yield such values. However, there may be many combinations of declared y_i's and λ parameters that yield the same utility values.

In practice, our extension introduces the $U^{*}(y)$ function into the $Z^{-1}(\ )$ density function, thereby adding a nonlinearity in (12), but implies the loss of only one degree of freedom, a practical consideration with functions estimated for each of a set of individuals.

Our generalization gets around the objection that Van Praag's index has equal intervals and is too rigid in assigning categories to utility levels (Ratchford, 1985)[3]. It does not modify the S-shape of the IWF (which, as pointed out by Kapteyn and Wansbeek (1985), violates Gossen's first law of diminishing marginal utility of increasing income for low values of y) or the fact that it is bounded at 1 and may only measure one's relative position in society.

An alternative use of our index, in its form (13), would consist in explaining utility directly by the socioeconomic variables used in Van Praag's last step. However, we would rather use our index in its form (1) because it is not sigmoïd: the estimated values of λ will decide whether Gossen's first law is violated ($\lambda > 0$) or not ($\lambda < 0$).

---

3 Ratchford points out that the highest level of n categories of utility is necessarily at (n-1)/n; in our case the maximum is arbitrarily set at 1 but the λ modifies the relative position of the highest level relative to 1.

It is an approach of this kind that we want to use, but with questionnaire answers that are cardinalized by the researcher. Our discussion and example need not be phrased in terms of utility: we simply assume that the level of an underlying phenomenon is of interest.

3.2 USEFULNESS WITH UNCARDINALIZED VALUES

Indeed, our index as described above assumes that the R_i values are provided directly by the answers. In many situations, this may not be the case: the answers contain ordered categories (excellent, good, etc.) but not numerical values. In these cases, the researcher can use ordered LOGIT and latent variable techniques. The first has the drawback of requiring as many models as questions considered. The second easily yields identification problems (Goldberger 1972). We want to suggest that our index is such that, given the willingness to substitute numerical values for the qualitative answers -a process we call cardinalization-, it provides a practical alternative to the use of ordered LOGIT and LISREL techniques.

Clearly, cardinalization by the researcher introduces "error". We will discuss the various sources of errors and will show their minimal effects on the imputed index and the results of the regression.

We argue that two principal sources of errors are involved when the researcher imputes numerical values to the ordinal results. The first concerns the length of the chosen domain; the second naturally involves inaccurate relative positions of the answers.

3.2.1 LENGTH OF THE DOMAIN

Consider first the effects on the SACOD index of a change in the length of the domain. For instance, what would be the consequences of a change in the original [1, .., 8] domain of R_i to the domain [1, .., 100]? If the minimal value of the domain is chosen to be one, we can formulate the question by the addition of one parameter to the definition to the index, namely

$$S = [\exp(\sum_{i=1}^{n} (R_i^{\gamma} - R_{max}^{\gamma})/n)]^{(\lambda)} \qquad (14)$$

The additional γ parameter[4] measures exactly a change in the length of the domain. In the next section, we will show with an empirical example that the SACOD index is relatively unaffected by significantly different values of the γ parameter: the λ parameter compensates for the changes in γ. Had the analyst associated numerical values from 1 to 100 to the ordinal answers this would have caused more or less the same results as numerical values from 1 to 8. This is a very interesting practical property. Note that this property is important as well for the determination of an arbitrarily chosen domain in the case of samples that contain cardinalized values.

3.2.2 INACCURACY OF THE RELATIVE POSITIONS OF THE ANSWERS

Now suppose that there are four possible answers to each question and that the researcher assigns the values (1, 2, 4, 8) to them. What would have been the consequences of choosing (1, 2.2, 3.8, 8) instead of (1, 2, 4, 8)? To handle this problem, we introduce a random term into the SACOD index as follows

$$S = [\exp(\sum_{i=1}^{n} (R_i^* - R_{max})/n)]^{(\lambda)}, \qquad (15)$$

where $R_i^* = R_i + \epsilon_i$ and ϵ_i is assumed to be zero if $R_i = R_{max}$ or R_{min} and to normally distributed with zero mean and constant variance otherwise.

This ϵ_i term modifies the relative importance of each possible answer. We argue that the values selected by the analyst, the chosen R_i, will not substantially affect the results.

The establishment of these two empirical properties is important since, when considered with the theoretical proposition already mentioned in Section 2, they mean that the analyst wishing to cardinalize ordinal answers has a significant latitude in assigning the numerical values to the R_i.

---

4 Note that one could not estimate an optimal γ because in this case the optional solution would be degenerate (i.e. $\gamma = 0$). Tests of γ can therefore only be made for purposes of numerical sensitivity analysis.

3.2.3 EXAMPLE: A GRIEVING REACTION INDEX

A. ILLUSTRATIVE MODEL

To illustrate the usefulness of the index, notably its self-adjustment property, and measure its sensitivity to cardinalization of data by the researcher, consider the following model[5] designed to explain the grieving reaction of a panel of individuals:

$$GR_{j,t} = f(P_j, T_j, D_j, S_j, N_j, Y_j, C_j, I_j, ET_j)_t + u_{j,t} \qquad (16)$$

where, for the j^{th} individual at time t, the dependent variable $GR_{j,t}$ denotes the value of the grieving reaction index, the capital letters on the right-hand side denote sets of explanatory variables and $u_{j,t}$ is a residual error of zero mean and constant variance.

The grieving reaction is defined by the answers to Goldberg's (1972) classic 30-item General Health Questionnaire (version for the United States), which uses a four-point response scale (less than usual/ same as usual/ rather more than usual/ much more than usual) to identify how an individual's present state differs from his usual state. The questions were answered in three interviews of a set of individual visitors selected at random in palliative care units of three large Montreal hospitals: the interviews were done at fixed intervals after the death of the visited patients. The numerical values that we associated with the four-point scale were in most cases 1, 2, 4 and 8. Form (1) of our index is used as dependent variable GR. We assume that the value of the index for each interviewed subject depends linearly upon P_j (the set of interviews the subject participated in), T_j (the passage of time), D_j (the deceased patient's characteristics and relationship to the subject), S_j, N_j and Y_j (the subject's sociological, environmental and economic characteristics), C_j (the support and professional help received by the subject), I_j (the interviewer's bias) and other factors ET_j.

---

5 This example uses a subsample and an exploratory submodel drawn from a larger study of bereavement that Louise L. Cuddihy is preparing with the present authors: in that study, the retained specification will no doubt differ somehow from that used in our illustrative example. The data are described in Kiely and Strobel (1985) who also present a quantitative analysis of several indexes of grieving built from the sample; this analysis of variance relates these indices to categories of subjects.

Parameter estimates were obtained by maximizing the likelihood of the cardinalized observations on dependent variable under the assumption of normality of the error term, using the L-1.2 algorithm (Liem et al., 1987). A grid search was done to insure that a global maximum had been found.

Table 1 contains, for each experiment performed, results presented in three sections. The first section lists elasticities of the index, defined for each explanatory variable X_h as

$$\frac{\partial GR}{\partial X_h} \cdot \frac{\overline{X}_h^+}{\overline{GR}} \qquad (17)$$

where $\overline{X}_h^+$ denotes the sample mean computed only with strictly positive observations[6]. Under each elasticity the *t*-statistic of the regression coefficient, computed conditionally upon the estimated value of λ, is shown in parentheses. The second section contains estimated values for λ and *t*-statistics (with respect to 0 and 1, successively). The third section contains general statistics, notably the log-likelihood and the probability that the sample contain observations that would violate the normality assumption of the error term.

As our objective here is not to discuss the specifics of the model results[7] but to illustrate the characteristics of our index with a real example, our comments will be selective.

---

6 As shown in Dagenais et al. (1987), this procedure includes both the standard definition of elasticity (for regressors that contain no observed zeroes), and the elasticity for dummy variables defined as the percentage change in the dependent variable due to the presence of X_h. In Table 1, variable code names that are underlined once or twice denote a dummy variable.

7 For a full exposition of the model, sample and results, the reader will consult the forthcoming paper mentioned above. To interpret the elasticities shown in Table 1, one should note that most variables denote mutually exclusive states. To avoid multicollinearity, one state (shown in parentheses at the end of the definition variable), is taken as reference state. A definition ending in MD denotes a missing data dummy variable.

TABLE 1

Determinants of grieving reaction:

Self adjustment and error sensitivity of the index

E L A S T I C I T I E S (T-STATISTICS)	CODE NO. = DEP. VAR. =	1 GR	2 GR	3 GR	4 GR	5 GR	6 GR
P = PARTICIPATION							
TIME OF INTERVIEWS 1-2 (1-2-3)	MOM12	.049	.005	.000	.039	.062	.057
	=====	(.91)	(.20)	(.01)	(.82)	(.98)	(1.00)
TIME OF INTERVIEWS 2-3 (1-2-3)	MOM23	.155	.052	.040	.131	.189	.170
	=====	(1.78)	(1.68)	(1.49)	(1.85)	(1.58)	(1.75)
TIME OF INTERVIEWS 1-3 (1-2-3)	MOM13	.233	.086	.065	.186	.304	.263
	=====	(1.90)	(2.04)	(1.77)	(1.91)	(1.70)	(1.86)
TIME OF INTERVIEW 1 (1-2-3)	MOM1	.140	.053	.042	.119	.169	.154
	====	(1.64)	(1.65)	(1.52)	(1.69)	(1.52)	(1.64)
TIME OF INTERVIEW 2 (1-2-3)	MOM2	.177	.050	.034	.159	.196	.185
	====	(1.81)	(1.22)	(.95)	(1.90)	(1.61)	(1.78)
TIME OF INTERVIEW 3 (1-2-3)	MOM3	1.083	.471	.372	.998	1.203	1.151
	====	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
T = TIME							
TIME BETWEEN DIAGNOSIS AND DEATH	TEMPSDIA	.004	.000	-.000	.003	.003	.004
	---------	(.24)	(.01)	(-.03)	(.25)	(.19)	(.27)
TIME BETWEEN DIAGNOSIS AND DEATH MD	TEMPSDIO	.033	.040	.038	.032	.038	.035
	=========	(.64)	(1.59)	(1.76)	(.68)	(.62)	(.63)
MONTHS SINCE DEATH	TEMPSDE	-.095	-.043	-.035	-.083	-.114	-.100
		(-1.94)	(-2.62)	(-2.51)	(-2.11)	(-1.65)	(-1.85)
D = DECEASED							
SPOUSE OF DECEASED (RELATIVE)	LIENCJT	.058	.043	.037	.036	.105	.078
	--------	(.57)	(.86)	(.87)	(.39)	(.87)	(.73)
PARENT OF DECEASED (RELATIVE)	LIENPRT	.214	.132	.109	.174	.292	.256
	--------	(1.14)	(1.64)	(1.59)	(1.05)	(1.22)	(1.23)
FRIEND OF DECEASED (RELATIVE)	LIENAMI	.247	.116	.096	.232	.265	.237
	--------	(1.45)	(1.48)	(1.40)	(1.55)	(1.28)	(1.33)
SON OR DAUGHTER OF DCD (RELATIVE)	LIENFLSE	.093	.049	.041	.078	.118	.103
	---------	(.94)	(1.01)	(.98)	(.89)	(.99)	(.98)
BROTHER OR SISTER OF DCD (RELATIVE)	LIENFRSR	.111	.054	.044	.092	.146	.122
	---------	(.93)	(.96)	(.91)	(.87)	(1.03)	(.98)
PATIENT AGE AT DEATH	AGEPAT	.082	.022	.013	.079	.074	.081
	------	(.60)	(.36)	(.25)	(.64)	(.49)	(.57)
PATIENT AGE AT DEATH MD	AGEPATO	.083	.026	.016	.075	.090	.087
	=======	(.58)	(.39)	(.27)	(.57)	(.57)	(.58)
S = SOCIOLOGICAL							
SEXE (FEMALE)	SEXE	-.153	-.067	-.054	-.132	-.183	-.166
	====	(-2.16)	(-3.07)	(-2.86)	(-2.41)	(-1.79)	(-2.07)
FRENCH SPEAKING CANADIAN (OTHER)	CANF	-.071	-.016	-.009	-.066	-.069	-.070
	----	(-.90)	(-.50)	(-.33)	(-.93)	(-.81)	(-.85)
ENGLISH SPEAKING CANADIAN (OTHER)	CANA	-.074	-.027	-.022	-.074	-.065	-.072
	----	(-1.23)	(-1.07)	(-.98)	(-1.38)	(-.96)	(-1.13)
CATHOLIC SUBJECT (OTHER)	RELCAT	-.207	-.092	-.073	-.165	-.287	-.239
	------	(-1.90)	(-2.09)	(-1.90)	(-1.88)	(-1.74)	(-1.89)
PROTESTANT SUBJECT (OTHER)	RELPROT	-.352	-.133	-.103	-.292	-.453	-.401
	-------	(-2.28)	(-2.89)	(-2.54)	(-2.57)	(-1.85)	(-2.16)
JEWISH SUBJECT (OTHER)	RELJUIF	-.058	.004	.011	-.037	-.098	-.084
	-------	(-.63)	(.10)	(.29)	(-.45)	(-.87)	(-.83)

RELIGION MD	RELO	-.294	-.102	-.074	-.242	-.376	-.337
	====	(-1.93)	(-1.77)	(-1.48)	(-2.00)	(-1.71)	(-1.92)
SUBJECT AGE AT DEATH	AGESUJ	.177	.081	.070	.185	.144	.167
	------	(1.09)	(1.17)	(1.19)	(1.27)	(.80)	(.99)
SUBJECT AGE AT DEATH MD	AGESUJO	.180	.081	.069	.185	.151	.170
	=======	(1.07)	(1.14)	(1.13)	(1.23)	(.82)	(.98)
MARRIED SUBJECT (SINGLE)	SMARIE	.123	.044	.034	.103	.147	.136
	------	(1.35)	(1.15)	(1.04)	(1.31)	(1.32)	(1.39)
SEPARATED SUBJECT (SINGLE)	SSEPARE	-.047	-.029	-.024	-.040	-.068	-.050
	-------	(-.38)	(-.44)	(-.42)	(-.36)	(-.47)	(-.39)
WIDOWED SUBJECT (SINGLE)	SVEUF	.200	.088	.069	.168	.261	.221
	-----	(1.44)	(1.58)	(1.44)	(1.43)	(1.44)	(1.46)
CHILDREN LESS THAN 14 YEARS	ENFANTS	.001	.002	.002	.001	.000	.001
	-------	(.08)	(.56)	(.65)	(.15)	(.05)	(.07)
CHILDREN LESS THAN 14 YEARS MD	ENFANTSO	.025	.030	.028	.026	.022	.015
	========	(.16)	(.36)	(.38)	(.18)	(.13)	(.09)
N = ENVIRONMENTAL							
DEPART. OF CHILDREN SINCE DEATH (NO)	DEPART	.087	.023	.017	.079	.094	.092
	======	(1.26)	(.84)	(.73)	(1.33)	(1.12)	(1.23)
MAJOR LOSSES SINCE DEATH (NO)	PERTES	.104	.057	.048	.090	.125	.111
	======	(1.86)	(3.18)	(3.12)	(2.03)	(1.62)	(1.78)
CONCURRENT CRISIS TO DEATH (NO)	CRISE	.119	.067	.058	.104	.139	.128
	=====	(1.98)	(3.09)	(3.10)	(2.20)	(1.61)	(1.90)
CONCURRENT CRISIS TO DEATH MD	CRISEO	-.075	-.031	-.025	-.065	-.090	-.078
	======	(-1.00)	(-.86)	(-.79)	(-.96)	(-1.03)	(-1.00)
Y = ECONOMIC							
FINANCIAL EFFECT; IMPROVEMENT (C)	FINA	-.006	.001	.002	-.007	-.006	-.006
	----	(-.12)	(.05)	(.08)	(-.15)	(-.09)	(-.11)
FINANCIAL EFFECT; DETERIORATION (C)	FIND	-.185	-.106	-.091	-.168	-.209	-.198
	----	(-2.45)	(-5.36)	(-5.40)	(-2.94)	(-1.87)	(-2.28)
SOCIO-EC. CATEGORY; PROFESS. (BC)	CSPWP	.089	.046	.038	.081	.098	.090
	-----	(.86)	(.95)	(.91)	(.86)	(.81)	(.82)
SOCIO-EC. CATEGORY; WHITE COLLAR(BC)	CSPWW	.157	.077	.064	.130	.200	.178
	-----	(1.49)	(1.75)	(1.69)	(1.45)	(1.43)	(1.53)
SOCIO-EC. CATEGORY; UNEMPLOYED (BC)	CSPUN	.424	.209	.173	.360	.523	.464
	-----	(1.99)	(3.01)	(2.97)	(2.05)	(1.83)	(1.97)
SUBJECT NOT WORKING (BC)	CSPNW	.006	.002	.000	-.008	.035	.021
	-----	(.09)	(.06)	(.01)	(-.13)	(.47)	(.31)
SOCIO-EC. CATEGORY MD	CSPO	-.050	-.026	-.022	-.056	-.033	-.038
	====	(-.42)	(-.44)	(-.43)	(-.53)	(-.24)	(-.31)
PRINCIPAL WAGE EARNER; PROFESS.(BC)	OCMWP	-.091	-.040	-.032	-.087	-.089	-.087
	-----	(-1.04)	(-1.00)	(-.91)	(-1.12)	(-.86)	(-.93)
PRINCIPAL WAGE EARNER; W. COLLAR (BC)	OCMWW	-.207	-.107	-.089	-.177	-.257	-.224
	-----	(-2.00)	(-3.05)	(-2.95)	(-2.18)	(-1.70)	(-1.91)
PRINCIPAL WAGE EARNER: UNEMPLOY.(BC)	OCMUN	-.105	-.034	-.025	-.079	-.156	-.126
	-----	(-.97)	(-.73)	(-.61)	(-.83)	(-1.13)	(-1.05)
PRINCIPAL WAGE EARNER; NOT WORK.(BC)	OCMNW	-.140	-.056	-.043	-.120	-.170	-.154
	-----	(-1.58)	(-1.66)	(-1.49)	(-1.64)	(-1.44)	(-1.56)
PRINCIPAL WAGE EARNER MD	OCMO	-.094	-.039	-.031	-.083	-.108	-.097
	====	(-1.12)	(-.95)	(-.88)	(-1.09)	(-1.08)	(-1.10)
C = CARE							
R.V.H. HOSPITAL (P.C.U.)	HOPRVH	-.118	-.049	-.039	-.107	-.134	-.124
	------	(-1.80)	(-2.12)	(-1.92)	(-1.99)	(-1.55)	(-1.72)
M.C.H. HOSPITAL (P.C.U.)	HOPMCH	-.077	-.033	-.025	-.073	-.084	-.079
	------	(-1.23)	(-1.13)	(-1.01)	(-1.31)	(-1.11)	(-1.15)
REFUSAL OF INTERVENTION (C)	REFUS	.025	.015	.012	.012	.054	.032
	=====	(.21)	(.25)	(.23)	(.11)	(.41)	(.26)
INTERVENTION (C)	INTERV	-.008	-.005	-.004	-.008	-.007	-.006
	======	(-.16)	(-.19)	(-.19)	(-.17)	(-.12)	(-.12)
INTERVENTION BY VOLUNTEER (NURSE)	BENEVOLE	-.068	-.042	-.037	-.059	-.083	-.073
	========	(-1.27)	(-1.89)	(-1.95)	(-1.29)	(-1.24)	(-1.26)

I = INTERVIEW							
INTERVIEWER 1 (17)	EVAL1	-.421	-.150	-.117	-.385	-.456	-.437
		(-2.11)	(-2.87)	(-2.57)	(-2.48)	(-1.73)	(-1.99)
INTERVIEWER 6 (17)	EVAL6	-.041	-.019	-.017	-.044	-.033	-.034
		(-.47)	(-.43)	(-.44)	(-.56)	(-.33)	(-.38)
INTERVIEWER 9 (17)	EVAL9	.347	.170	.145	.340	.334	.348
		(3.10)	(3.58)	(3.59)	(2.52)	(1.61)	(1.95)
INTERVIEWER 21 (17)	EVAL21	-.141	-.081	-.069	-.134	-.140	-.143
		(-1.23)	(-1.16)	(-1.08)	(-1.32)	(-1.06)	(-1.17)
INTERVIEWER 10 (17)	EVAL10	-.060	-.012	-.009	-.051	-.063	-.072
		(-.64)	(-.26)	(-.23)	(-.60)	(-.58)	(-.71)
INTERVIEWER 12 (17)	EVAL12	-.162	-.052	-.040	-.155	-.159	-.165
		(-1.26)	(-.79)	(-.66)	(-1.39)	(-1.02)	(-1.19)
INTERVIEWER 14 (17)	EVAL14	-.128	-.052	-.041	-.116	-.136	-.132
		(-1.19)	(-1.07)	(-.97)	(-1.24)	(-1.05)	(-1.14)
INTERVIEWER 15 (17)	EVAL15	-.226	-.067	-.048	-.199	-.254	-.243
		(-1.86)	(-1.46)	(-1.21)	(-2.01)	(-1.57)	(-1.78)
INTERVIEWER 16 (17)	EVAL16	.143	.041	.029	.156	.099	.132
		(1.17)	(.75)	(.60)	(1.42)	(.74)	(1.04)
INTERVIEWER 18 (17)	EVAL18	-.058	-.016	-.011	-.043	-.083	-.080
		(-.60)	(-.34)	(-.27)	(-.50)	(-.72)	(-.77)
INTERVIEWER 2 (17)	EVAL2	.060	.011	.004	.039	.101	.071
		(.59)	(.22)	(.10)	(.42)	(.87)	(.67)
INTERVIEWER 3 (17)	EVAL3	.069	.052	.044	.061	.096	.064
		(.71)	(1.12)	(1.11)	(.70)	(.85)	(.63)
INTERVIEWER 4 (17)	EVAL4	.032	.029	.024	.015	.075	.044
		(.21)	(.39)	(.37)	(.11)	(.45)	(.28)
INTERVIEWER 5 (17)	EVAL5	-.036	-.028	-.029	-.068	.048	-.016
		(-.30)	(-.51)	(-.60)	(-.63)	(.36)	(-.13)
INTERVIEWER 7 (17)	EVAL7	.125	.037	.026	.115	.147	.128
		(.93)	(.64)	(.54)	(.94)	(.98)	(.91)
INTERVIEWER 19 (17)	EVAL19	-.010	.000	.001	-.013	.003	-.009
		(-.09)	(.01)	(.02)	(-.13)	(.02)	(-.08)
INTERVIEWER 8 (17)	EVAL8	.290	.130	.105	.297	.260	.272
		(1.37)	(1.64)	(1.54)	(1.54)	(1.13)	(1.25)
INTERVIEWER 20 (17)	EVAL20	.247	.096	.070	.220	.302	.268
		(.83)	(.83)	(.71)	(.83)	(.88)	(.84)
INTERVIEWER 13 (17)	EVAL13	.109	.033	.024	.124	.059	.088
		(.85)	(.58)	(.48)	(1.04)	(.43)	(.67)
ET = ET CETERA							
COPING PREDICTION	COPING	-.213	-.124	-.109	-.193	-.247	-.226
		(-2.24)	(-3.90)	(-3.94)	(-2.58)	(-1.75)	(-2.09)
COPING PREDICTION MD	COPINGD	-.176	-.118	-.106	-.162	-.199	-.183
		(-1.79)	(-2.97)	(-3.05)	(-1.98)	(-1.49)	(-1.70)
CONSTANT	CONSTANT	-.184	-.492	-.852	-.060	-3.110	-.165
		(-2.72)	(-8.94)	(-10.45)	(-3.38)	(-2.01)	(-2.48)
P A R A M E T E R S	CODE NO. =	1	2	3	4	5	6
(T-STATISTICS)	DEP. VAR. =	GR	GR	GR	GR	GR	GR
BOX-COX TRANSFORMATIONS							
LAMBDA(Y)	LAM Y	-5.202	0.000	1.000	-13.359	-.996	-5.973
		(-7.91)	FIXED	FIXED	(-6.27)	(-11.34)	(-8.62)
		(-9.43)			(-6.74)	(-22.72)	(-10.06)
LOG-LIKELIHOOD		1037.286	930.223	875.573	1498.446	2687.383	1066.583
PSEUDO-R2		.656	.499	.393	.518	.997	.693
NUMBER OF OBSERVATIONS		570	570	570	570	570	570
NUMBER OF INDEPENDENT VARIABLES		73	73	73	73	73	73
AVERAGE PROBABILITY (Y=LIMIT OBSERVATION)		0.000	0.000	0.000	0.000	0.000	0.000

B. SELF ADJUSTMENT AND SENSITIVITY TO CARDINALIZATION

The first three columns of Table 1 clearly show the benefit from using the Box-Cox transformation: the optimal model in Column 1 indicates an optimal model with a clearly concave form because λ is negative and significantly different from zero. The considerable drop in likelihood between Column 1 and Column 2 indicates that the linear arithmetic mean model is clearly inferior to the self-adjusted model: this is of interest because researchers sometimes add the scores obtained on different items to obtain an overall score: that is close to setting the λ at zero in our model[8]. Similarly, the geometric mean model implied by setting λ=1 is rejected even more strongly than the arithmetic mean model, as the drop in likelihood shows in Column 3 indicates. The optimal model of the first column implies diminishing marginal contributions to grieving of higher scores on any of the 30 items.[9]

The last three columns show the sensitivity of all results to the cardinalization procedure. In contrast with Columns 1-3 and 6, where the γ parameter determining the size of the domain is set arbitrarily at 1, it is set at .5 and 2, respectively, in Column 4 and Column 5. One sees in the second section of the table that modifications of the power γ are essentially compensated by modification of the λ; in the first section, it is clear that the elasticities are not very affected by the modifications of γ. The sixth column shows the effect of generating errors in relative positions of the observations: the R_i values were modified by addition of a random term, distributed normally with a standard error of .4 (i.e. about 10% of the mean value of R_i), to the specification shown in Column 1: the effect on the elasticities, *t*-statistics and other parameters is very small.

---

8 Our procedure is not a simple addition of scores, but of deviations from the maximum score, in that case.

9 Because the estimated value of λ differs from 0, the results also imply that there is an interaction among the different dimensions or components of grieving included in the index, as is clear from the product formulation (1-A). This interaction should not be confused with that found in an altogether different index that would use a distinct λ_i for each component.

4.CONCLUSION

We have defined a self-adjusting index to account for phenomena measured through the use of questionnaires. We have analyzed the formal properties of the index under the assumption that questionnaire answers were cardinal and shown, with an example, that it could also be used with ordinal answers cardinalized by the researcher. With cardinal answers the index has the following properties: data scale compensation, indifference to shift of analyst's values, indifference to the location of the origin and endogenous curvature. The empirical example leads us to believe that the self-adjustment property of the index yields results that strongly dominate more constrained specifications and that the index is not very sensitive to the length of the domain of cardinalization or to a 10% inaccuracy in relative positions of the answers. These last two properties suggest that the index is relevant not only as a way of synthesizing cardinal answers but also as a novel way of extracting information from data sets containing ordinal choices made by individuals.

5. REFERENCES

Dagenais M.G., M.J.I. Gaudry and T.C. Liem, 1987, "Urban Travel Demand: An Impact of Box-Cox Transformation with Nonspherical Residual Error", Transportation Research B, 21, 6, 443-477.

Gaudry M.J.I., and R. Laferrière, 1988, "The Box-Cox Transformation: Power Invariance and a New Interpretation", Publication #573, Centre de recherche sur les transports, Université de Montréal, 1988.

Goldberg D.P., 1972, The Detection of Psychiatric Illness by Questionnaire, Institute of Psychiatry, Maudsley Monographs, Oxford University Press.

Goldberger A., 1972, "Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables",International Economic Review 13, 1-15.

Kapteyn A. and T. Wansbeek, 1985, "The Individual Welfare Function", Journal of Economic Psychology, 6, 333-363.

Kiely M.C. and M.G. Strobel, 1985, "Evaluation of a Bereavement Intervention Program", Département de psychologie, Université de Montréal.

Ratchford B.T., 1985, "The Individual Welfare Function: A Comment", Journal of Economic Psychology, 6, 365-374.

Schlesselmann J., 1971, "Power Families: A Note on the Box and Cox Transformation", Journal of the Royal Statistical Society B 33, 307-311.

Spitzer J.J., 1984, "Variance Estimates in Models with the Box-Cox Transformation: Implications for Estimation and Hypothesis Testing", The Review of Economics and Statistics 66, 645-652.

Van Praag B., 1968, Individual welfare function and consumer behavior, Contributions to Economic Analysis, North Holland, Amsterdam.